**Mendelian Randomization Study**

**Abstract**

**Context:** Most types of chronic liver diseases (CLDs) are associated with type 2 diabetes (T2D). However, it is uncertain whether causality exists between CLDs and T2D.

**Objective:** To confirm whether two main types of CLDs nonalcoholic fatty liver disease (NAFLD) and hepatitis C virus (HCV), can lead to T2D.

**Design and methods:** We selected independent genetic variants of NAFLD (272 cases, 96,227 controls) and HCV (5,794 cases,206,659 controls) based on published genome-wide association studies (GWAS) as instrumental variables (IVs). And we performed a two-sample Mendelian randomization (MR) analysis on T2D (62,892 cases, 596,424 controls) to explore their relationship.

**Results:** The MR results showed that there were high ORs (1.0770, 95%CI 1.0599 to 1.0943, p = 9.10E-20 for NAFLD to T2D; 1.1434,95%CI 1.0741 to 1.2172, p=2.65E-05 for HCV to T2D) and low intercepts, which proved the causality between the two main types of CLDs and T2D.

**Conclusions:** Our study shows that two main types of CLDs, NAFLD and HCV, can contribute to T2D. Both NAFLD and HCV can lead to insulin resistance (IR). Our study may provide an evidence for a model that CLDs can lead to T2D through IR.

**Key words:** Mendelian randomization, CLD, T2D, causality, GWAS, SNPs

# Introduction

Diabetes mellitus (DM) is a chronic metabolic disorder, which has a steadily increasing prevalence for decades. There were approximately 463 million DM patients aged 20 to 79 years old until 2019, and it is estimated to be 578 million DM patients in 2030 and 700 million in 2045[1](#_ENREF_1). In addition, diabetes can lead to several microvascular complications and macrovascular complications which can result in blindness, renal failure, myocardial infarction and stroke[2](#_ENREF_2).

The liver plays an important role in the body metabolic processes, especially in glucose homeostasis. CLDs such as NAFLD and HCV are associated with insulin resistance, which indicate the probable relation between CLDs and T2D[3](#_ENREF_3),[4](#_ENREF_4). PatientswithNAFLD have a high risk of T2D and vice versa[5](#_ENREF_5),[6](#_ENREF_6). It seems that they have a bi-directional relationship, but the causality between them is still unconfirmed[7](#_ENREF_7),[8](#_ENREF_8). Similarly, HCV infection represents a well-known risk factor for T2D[9](#_ENREF_9). However, on account of the influence of confounding factors, we could not draw conclusion that there is a causality between CLDs and T2D directly. In addition, it is difficult to judge the sequence of exposure and outcome, which can lead to reverse causality.

In this study, we used two-sample MR to determine the potential causality between CLDs and T2D. MR uses genetic data to probe questions of causality in epidemiological research, usually single nucleotide polymorphisms (SNPs), as instrumental variables (IVs) in epidemiologic study[10](#_ENREF_10),[11](#_ENREF_11). Each person obtains an allele at the SNP locus before any confounders occur, which means no confounding or reverse causation. Thus, our analysis provides a better understanding of the pathogenic mechanisms and potential correlation between these human complex genetic diseases.

# Materials and methods

Two-sample MR uses publicly available summary-level data from GWAS, which has three main assumptions: (a) IVs should be associated with the exposure, and herein we just select SNPs of genome wide significant association (p<5E-08); (b) IVs should not be associated with confounding factors; (c) IVs should influence the outcome only through the exposure(Figure 1). Therefore, we should consider potential violations of the IV assumptions such as instruments affect the outcomes directly, or through other ways (not affect the exposures) affect the outcomes. In other word, we should estimate the presence of horizontal pleiotropy, inheritance and linkage disequilibrium (LD).



Figure 1. A directed acyclic graph of MR model.

Assumption a: IVs are correlated with exposure; assumption b: IVs affect outcomes only through exposure; assumption c: IVs are not related to confounding factors.

## Instrument variable selection

We selected SNPs associated with T2D from a GWAS study published by Angli et al., which involved 62,892 T2D cases and 596,424 control subjects[13](#_ENREF_13). The HCV dataset was extracted from a GWAS study by Kazuyoshi et al. including 5,794 cases and 206,659 control subjects[14](#_ENREF_14). And we obtained SNPs associated with NAFLD including 272 cases and 96,227 control subjects benefiting from FinnGen study. FinnGen is a public-private research project covering the whole of Finland, aimed at making breakthroughs in disease prevention and diagnosis.

To satisfy three IV assumptions above, we had a series of steps to select eligible SNPs from the GWAS. We only selected SNPs associated with CLDs or T2D at a genome-wide association (p < 5E−08). And genetic variants without independent inheritance (r2> 0.01) and with potential LD were removed.

## MR analysis and Sensitivity analysis

Inverse variance weighted (IVW) was used as the key method to calculate the causal effect between CLDs and T2D for Two-sample MR[15](#_ENREF_15). IVW method is calculated as follows:

where represents the association effect of IVs on T2D, defines the association effect of IVs on CLDs, and represents the IVW estimate which is calculated as and is the residual term.

We performed several sensitivity analyses to exclude the impact of heterogeneity and pleiotropy within the genetic instruments, which can bias the MR results. MR-Egger was performed to assess and adjust for horizontal pleiotropy[16](#_ENREF_16):

where represents the effect between IVs and T2D; represents the intercept term, represents the MR-Egger estimate, and represents the residual term. The intercept can represent the estimated average value of horizontal pleiotropic. Then we conducted the Cochran Q statistics and I2 statistics, which represent the size of heterogeneity. The Cochran Q statistic was calculated as the weighted sum of the squared differences between individual SNP effects and the pooled effect across all SNPs. Weighted median estimator, penalized weight median, simple mode, and weight mode were performed to account for potential violations of the IV assumptions, presence of horizontal pleiotropy, heterogeneity, and error in the instrument-exposure associations[17](#_ENREF_17).

In addition, we conducted the global test implemented in the “MR-PRESSO” R package to evaluate the presence of pleiotropy among the genetic instruments[18](#_ENREF_18). If the global test was significant (p-value < 0.05), we removed outlying genetic variants (p-value < 0.05) and performed the IVW analysis again. MR analysis were performed with “TwoSampleMR”, “MendelianRandomization”, “MRPRESSO”, and “Meta” packages in R version 3.5.3.

# Results

## The influence of chronic liver diseases on the risk of type 2 diabetes

We extracted 4 genome-wide significant SNPs of NAFLD and 5 SNPs of HCV, which excluded pleiotropic SNPs identified statistically by MR-PRESSO (Table 1, Table 2). By using IVW analysis, we obtained ORs (1.0770; 95%CI 1.0599 to 1.0943; p = 9.10E-20) for NAFLD on T2D and ORs (1.1434; 95%CI 1.0741 to 1.2172; p=2.65E-05) for HCV on T2D, respectively. The results showed that NAFLD and HCV are considerably related to the high risk of type 2 diabetes separately. Since there was no significant heterogeneity between genetic variants (p= 0.99 for NAFLD to T2D; p=0.23 for HCV to T2D), we chose fixed model for MR analysis. In addition, we used 3 other methods to analyze, including MR Egger, Weighted median, and MR-PRESSO test, and found similar results (Table 3).

## Sensitivity analysis

Leave-one-out cross validation was used to estimate whether a single SNP can affect the MR results. For both NAFLD to T2D and HCV to T2D, the plot results from leave-one-out analysis showed that no matter which genetic variants was removed, the β value was always located to the right of zero, which indicated every point had the same direction effect on the result, and there was no significant offset. Furthermore, we made forest plots to evaluate the heterogeneity of these genetic variants, and we calculated the Cochran Q statistics and I2 statistics(Figure 2)(Figure 3). All the results indicated that there were no sensitive genetic variants among the IVs.



Figure 2. Forest plot of NAFLD on T2D. I2 and p (Cochran Q statistics) indicate the size of heterogeneity. Low I2 and high p means little heterogeneity exists. I2 =0% and p=1 indicate there is almost no heterogeneity.



Figure 3. Forest plot of HCV on T2D.I2=46% and p=0.09 indicate there is slight heterogeneity.

# Discussion

In this study, we designed a two-sample Mendelian randomization study to explore the causal relationship between Chronic liver diseases and type 2 diabetes. We found high ORs (1.0770, 95%CI 1.0599 to 1.0943, p = 9.10Ee-20 for NAFLD to T2D; 1.1434,95%CI 1.0741 to 1.2172, p=2.65E-05 for HCV to T2D) by using IVW method. It provided potentially credible evidence that some CLDs likely increase the risk of T2D.

In previous studies, researchers have discovered the correlation between CLDs and T2D, which is supported and mediated by insulin resistance (IR). It is induced both by the direct effects of the HCV proteins, and by the indirect effects mediated by chronic inflammation, oxidative stress and hepatic steatosis[3](#_ENREF_3). Besides, IR also mediate the association between NAFLD and T2D[4](#_ENREF_4). HCV promotes IR status, if it lasts for a long time, pancreatic beta-cell disfunction occurs, ultimately leading to irreversible damage and development of T2D[19](#_ENREF_19). The process of NAFLD leading to T2D has similar pathophysiology mechanism, while it has some differences. NAFLD patients can regulate peripheral insulin concentration to compensate the increased muscle IR while most of CHC patients have an impairment in insulin response to oral glucose tolerance test which increases their risk to developing T2D[20](#_ENREF_20). Therefore, we speculate that CLDs may cause T2D through IR, although the mechanism of inducing IR may be different. There has been a study on confirming the causality between NAFLD and T2D using MR method[21](#_ENREF_21), but the causality between HCV and T2D is still not confirmed. And the causal relation between CLDs and T2D is also unexplored overall.

Our study benefits from both GWAS and MR method. It is noteworthy that heterogeneity among causal estimates derived from multiple genetic variations may violate the necessary IV assumptions. In our study, we obtained values of I2 and p in the forest plot (I2=0%, p=1 for NAFLD to T2D; I2=46%, p=0.09 for HCV to T2D), which indicated that there was no heterogeneity for causal estimate of NAFLD to T2D. But for HCV to T2D, though the values (I2=46%<50%, p=0.09>0.05) were acceptable[22](#_ENREF_22), they showed that slight heterogeneity existed. Meanwhile, due to the lack of relevant data, we didn’t explore the effect of type 2 diabetes to Chronic liver diseases.

To the best of our knowledge, it is the first time to reveal the causality between HCV and T2D. We confirmed the causal relationship between most CLDs and T2D. It supports a model that CLDs can lead to T2D through IR. It helps researchers to understand the relationship between CLDs and diabetes better from the perspective of wholeness. In addition, it may give a theoretical foundation to a new strategy for better management of diabetes. It has found that effective control of cirrhosis might be a good way of better management of diabetes and may be more effective if using different therapeutic methods to different etiologies of liver cirrhosis[23](#_ENREF_23).

# Acknowledgments

Summary-level data used in this study was downloaded from the GWAS Catalog (https://www.ebi.ac.uk/gwas/).The authors thank participants of FinnGen study and all investigators for sharing these data.

1. Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9(th) edition. *Diabetes Res Clin Pract.* 2019;157:107843.

2. Kanter JE, Bornfeldt KE. Impact of Diabetes Mellitus. *Arterioscler Thromb Vasc Biol.* 2016;36(6):1049-1053.

3. Nevola R, Acierno C, Pafundi PC, Adinolfi LE. Chronic hepatitis C infection induces cardiovascular disease and type 2 diabetes: mechanisms and management. *Minerva Med.* 2020.

4. Tanase DM, Gosav EM, Costea CF, et al. The Intricate Relationship between Type 2 Diabetes Mellitus (T2DM), Insulin Resistance (IR), and Nonalcoholic Fatty Liver Disease (NAFLD). *J Diabetes Res.* 2020;2020:3920196.

5. Yi M, Chen RP, Yang R, Chen H. Increased prevalence and risk of non-alcoholic fatty liver disease in overweight and obese patients with Type 2 diabetes in South China. *Diabet Med.* 2017;34(4):505-513.

6. Amiri Dash Atan N, Koushki M, Motedayen M, et al. Type 2 diabetes mellitus and non-alcoholic fatty liver disease: a systematic review and meta-analysis. *Gastroenterol Hepatol Bed Bench.* 2017;10(Suppl1):S1-s7.

7. Lonardo A, Nascimbeni F, Mantovani A, Targher G. Hypertension, diabetes, atherosclerosis and NASH: Cause or consequence? *J Hepatol.* 2018;68(2):335-352.

8. Mantovani A, Byrne CD, Bonora E, Targher G. Nonalcoholic Fatty Liver Disease and Risk of Incident Type 2 Diabetes: A Meta-analysis. *Diabetes Care.* 2018;41(2):372-382.

9. Negro F, Forton D, Craxì A, Sulkowski MS, Feld JJ, Manns MP. Extrahepatic morbidity and mortality of chronic hepatitis C. *Gastroenterology.* 2015;149(6):1345-1360.

10. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet.* 2014;23(R1):R89-98.

11. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med.* 2017;36(11):1783-1802.

12. Figure1.

13. Xue A, Wu Y, Zhu Z, et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun.* 2018;9(1):2941.

14. Ishigaki K, Akiyama M, Kanai M, et al. Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat Genet.* 2020;52(7):669-679.

15. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol.* 2013;37(7):658-665.

16. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol.* 2015;44(2):512-525.

17. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol.* 2016;40(4):304-314.

18. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet.* 2018;50(5):693-698.

19. Ballestri S, Nascimbeni F, Romagnoli D, Baldelli E, Targher G, Lonardo A. Type 2 Diabetes in Non-Alcoholic Fatty Liver Disease and Hepatitis C Virus Infection--Liver: The "Musketeer" in the Spotlight. *Int J Mol Sci.* 2016;17(3):355.

20. Svegliati-Baroni G, Gaggini M, Carli F, et al. Mechanisms for increased risk of diabetes in chronic liver diseases. *Liver Int.* 2020;40(10):2489-2499.

21. De Silva NMG, Borges MC, Hingorani AD, et al. Liver Function and Risk of Type 2 Diabetes: Bidirectional Mendelian Randomization Study. *Diabetes.* 2019;68(8):1681-1691.

22. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *Bmj.* 2003;327(7414):557-560.

23. Zhao Y, Xing H, Wang X, et al. Management of Diabetes Mellitus in Patients with Chronic Liver Diseases. *J Diabetes Res.* 2019;2019:6430486.

**TABLE 1 |** Associations of genetic variants between NAFLD and T2D

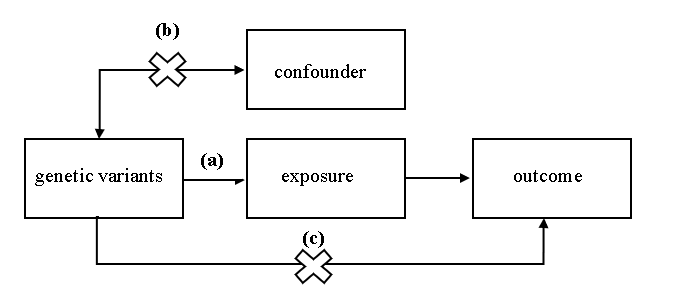
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **SNP** | **beta.exposure** | **se.exposure** | **pval.exposure** | **beta.outcome** | **se.outcome** | **pval.outcome** |
| rs11090617 | 0.636 | 0.1145 | 2.78E-08 | 0.1585 | 0.0105 | 6.38E-06 |
| rs12484700 | -0.6375 | 0.1132 | 1.80E-08 | 0.8410 | 0.0106 | 4.08E-06 |
| rs4823173 | -0.6361 | 0.1145 | 2.76E-08 | 0.8418 | 0.0106 | 4.15E-06 |
| rs738409 | 0.6061 | 0.107 | 1.49E-08 | 0.2182 | 0.0094 | 7.11E-06 |

**TABLE 2 |** Associations of genetic variants between HCV and T2D

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **SNP** | **beta.exposure** | **se.exposure** | **pval.exposure** | **beta.outcome** | **se.outcome** | **pval.outcome** |
| rs1042148 | 0.1494 | 0.02733 | 4.549E-08 | 0.0243 | 0.0090 | 0.0066 |
| rs2523608 | 0.1139 | 0.01955 | 5.633E-08 | -0.0022 | 0.0080 | 0.7845 |
| rs3094199 | 0.1502 | 0.02734 | 3.901E-08 | 0.0241 | 0.0089 | 0.0067 |
| rs3130981 | 0.1497 | 0.02733 | 4.311E-08 | 0.0231 | 0.0082 | 0.0047 |
| rs3132555 | 0.1493 | 0.02733 | 4.650E-08 | 0.0248 | 0.0089 | 0.0053 |

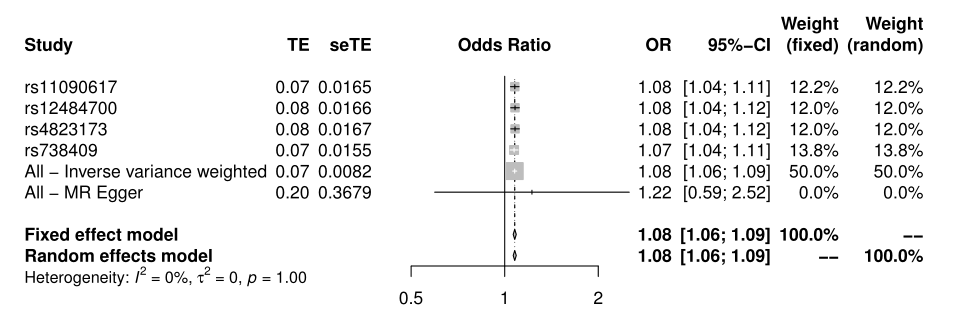
**TABLE 3 |** MR estimates from each method of assessing the causal effects of NAFLD and HCV on T2D risk

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Exposure traits** | **MR methods** | **Type 2 Diabetes** | | |  |
|  |  | **OR (95%CI)** | **SE** | **MR**  **P-value** | **P-value of Global Test in MR-PRESSO** |
| NAFLD | MR Egger | 1.2235(0.5948,2.5166) | 0.367944 | 0.638461 | 0.993 |
| Weighted median | 1.0783(1.0531,1.1042) | 0.012075 | 4.11E-10 |
| Inverse variance weighted | 1.077(1.0599,1.0943) | 0.008153 | 9.10E-20 |
| HCV | MR Egger | 2.0800(1.2628,3.426) | 0.254599 | 0.0637 | 0.388 |
| Weighted median | 1.1736(1.0913,1.2622) | 0.037113 | 1.60E-05 |
| Inverse variance weighted | 1.1434(1.0741,1.2172) | 0.031909 | 2.65E-05 |

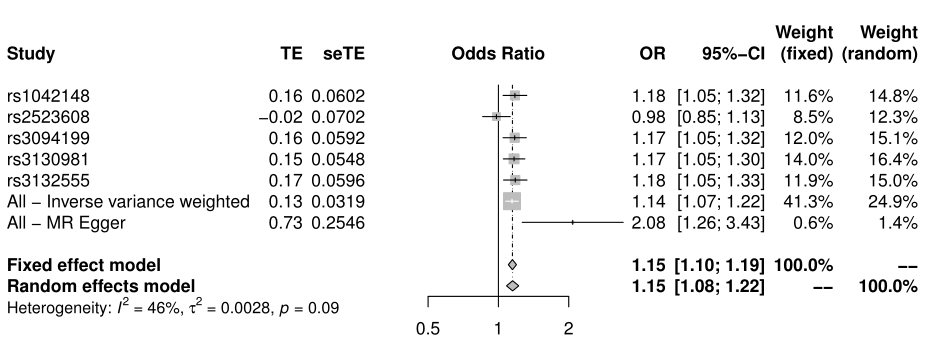


**FIGURE 1 |** A directed acyclic graph of MR model.

Assumption a: IVs are correlated with exposure; assumption b: IVs affect outcomes only through exposure; assumption c: IVs are not related to confounding factors.



**FIGURE 2 |** Forest plot of NAFLD on T2D. I2 and p (Cochran Q statistics) indicate the size of heterogeneity. Low I2 and high p means little heterogeneity exists. I2 =0% and p=1 indicate there is almost no heterogeneity.



**FIGURE 3 |** Forest plot of HCV on T2D.I2=46% and p=0.09 indicate there is slight heterogeneity.